勿以恶小而为之：滑坡效应的心理机制及预防策略

欧红蕾[1] 孙炳海[1,2] 张文海[3] 李伟健[1]

（[1]浙江师范大学教师教育学院；[2]浙江师范大学田家炳德育研究中心，金华 321004）

（[3]江苏省盐城工学院心理健康教育中心，盐城 224003）

摘 要 滑坡效应是指个体的不道德程度有随时间推移而逐渐加剧的倾向，即不道德行为具有反复性和渐进性。它广泛存在于个体自身层面以及人际互动过程中，会影响个体正常社会交往，甚至阻碍社会和谐健康发展。滑坡效应的研究方式包括实验者操纵法和自发性欺骗法，心理机制包括道德推脱、内隐偏差、自我非人化等，对应的干预措施有提升预防定向动机和塑造积极的描述性社会规范。未来研究可从心理过程和生理机制的角度进一步探讨滑坡效应的适用范围、神经干预方式等。

关键词 滑坡效应；道德推脱；内隐偏差；自我非人化；预防定向；描述性社会规范

分类号 B841

引言

三国时期，刘备在遗诏中告诫后人：“勿以恶小而为之”，以此劝勉其子刘禅要进德修业，勿行小恶，因为积小恶“足以乱国家”。南北朝时期，《后汉书》提出防微杜渐的思想，认为刚出现错误的思想或行为时，就应及时制止，避免其继续发展。自古以来，类似这种防患于未然的思想流传已久，预防不道德行为加剧的观念也早已深入人心。西方研究者将这种“不道德行为有重复和加剧的倾向”命名为滑坡效应(slippery slope)，即当个体在实施了不道德行为之后，其道德水平将如同雪球从滑坡上滚落，并难以抑制地急剧下降。

早在 1975 年，滑坡效应便在 Milgram 广为人知的服从实验中得以体现：被试在惩罚指令下对无辜者逐步施加了越来越强的电击，最后竟高达致命的程度(Milgram, 1975)。当时研究仅聚焦于权威服从，未有人发现其隐含的滑坡效应。此后，有研究者认为，Milgram 创造的这一实验情境之所以会导致高服从率，其中一个关键特征是电击强度的递增性质(Burger, 2009)。假设电击指令由最初的微弱电压突然提高到致命电压，也许就会有更多的被试拒绝执行这一伤害行为。

近年来，滑坡效应开始进入实证研究阶段，国内外研究者对滑坡效应开展了深入研究，在探索其心理、生理机制以及预防措施方面不断取得新的进展(陈银飞, 2011; Garrett, Lazzaro,

收稿日期：2019-07-02

*浙江省哲学社会科学新兴（交叉）学科重大扶持项目(ZZ323206060219015005)支持。

通信作者：李伟健，E-mail: xlxh@zjnu.cn

Ariely, & Sharot, 2016; Tang et al., 2019; Welsh, Ordonez, Snyder, & Christian, 2015; Yin & Weber, 2019)。本文尝试对滑坡效应进行系统、全面的综合阐述，倡导道德研究从停留于单个时间点转变为时间动态发展视角。本文首先对滑坡效应进行概述，其次总结其研究方法、成因以及干预方式，最后提出研究展望，以期为未来相关研究与应用提供一定参考。

1 滑坡效应的概念及研究概述

滑坡效应指个体的不道德行为往往存在反复性，且不道德程度会随时间推移而逐渐加剧。滑坡效应类似于小恶逐渐积累成大恶的过程，强调不道德行为的重复性和渐进性(Baron, Zhao & Miao, 2015; Lafollette, 2005; Welsh et al., 2015)。

早期的滑坡效应多为根据生活经验、社会热点和前人观点的理论型综述。如我国古训“小洞不补，大洞叫苦”、“防微杜渐”提醒人们需警惕坏事的苗头。国外学者 Schauer 根据大量真实法律案件得出结论：某个起初看起来没有违背道德标准的不良行为，却可能在接下来导致一连串危害性越来越大的相似事件发生(Schauer, 1985)；Lafollette 认为如果发生了不道德程度较小的 X，经过一系列微小的变动，就很可能会发生不道德程度极高的 Y(Lafollette, 2005)。这些观点都认为滑坡效应就如同温水煮青蛙，人难以注意自己的不符合伦理规范行为的缓慢变化，极容易酿成大祸。

近期的滑坡效应被国内外研究者引入了实证研究领域。国内研究中，郑琴琴(2009)采用事件史分析法，借助 2000~2007 年的纵向数据，考察了上市公司的信息披露违规行为，结果发现企业前期的不道德行为和后续不道德行为的发生率之间存在高度相关。陈银飞和茅宁(2011)运用实验法，研究了企业决策者的伦理判断偏差的时间变化过程，结果表明不符合伦理规范的行为是逐渐变严重的，而非突然加剧的。此后国内研究未在该领域有太多进展。与此同时，国外研究表明，在奖励逐渐增多而非突然增多的情况下，个人更有可能从事不道德行为(Welsh et al., 2015)，接纳他人的不道德行为(Gino & Bazerman, 2009)。不道德行为的性质会随着时间变化越来越严重(Engelmann & Fehr, 2016)。近期国内学者开始继续跟进，利用近红外光谱技术、功能磁共振成像等技术对滑坡效应的神经机制展开了进一步探讨(Tang et al., 2019; Yin et al., 2019)。

总结以往研究，滑坡效应存在于个体自身层面和人际互动层面。个体自身层面，指不道德行为的滑坡发生于个体自身，未对他人产生影响的情况。如 Welsh 等人证明了个体层面滑坡效应的存在，被试的欺骗量随时间逐渐增多(Welsh et al., 2015)。人际互动层面，指在人与人、群体与群体之间的互动中也容易出现滑坡效应(Ashforth & Anand, 2003; Courtois &

Gendron, 2017)。如 Gino 等人研究揭示了人际层面的滑坡效应，若被试双方互为合作者，其中一个人做出严重性递增的不道德行为，另一个人也会逐渐参与其中(Gino et al., 2009)。Volokh 提出群际层面的滑坡效应(Volokh, 2003)，认为滑坡效应容易发生在群际。如群体极化现象(group polarization)，一部分群体成员最初的某些偏差观点或行为最终可能会导致整个群体层面的暴行(Moscovici & Zavalloni, 1969)。

在提及滑坡效应时，要警惕“滑坡谬论”。滑坡谬论指不合理地使用一连串的因果关系，将所有的“可能性”转换成“必然性”，以得出某些人妄图证实的因果结论。滑坡谬论的逻辑是：如果发生 A，那就一定会发生 B，接着引发 C……最终导致严重后果 Z 的发生。因此，绝对不能允许 A 的发生。这种“如果某某事件发生，那么某种坏事会随之发生”的说法过于极端和片面，在滑坡效应的理解上，切忌一概而论，须知滑坡效应只是代表一种倾向性，不能将其解释为必然性。

总之，不道德行为存在随时间逐渐变化的倾向，这种倾向会把人的不道德行为推向愈加败坏的地步，这个过程被称为滑坡效应。根据其特点，下一节将介绍滑坡效应的研究方法。

2 滑坡效应的研究方法

基于滑坡效应的重复性和渐进性，滑坡效应的研究方法有其独特之处：只要对道德的经典范式稍加改编，就能很好地模拟单人或人际互动模式中的滑坡效应。当前已被改编用于滑坡效应研究的方式有两大类：一是实验者操纵法，该方式需要实验者额外操纵出滑坡组和突变组；二是自发性欺骗法，这不需要额外操纵，仅将轮数延长来观察被试行为的自发性变化即可。

2.1 实验者操纵法

该方法产生于滑坡效应的早期研究中(陈银飞, 2011)，通常从任务奖励或难度入手，将相关变量操纵为“渐变”和“突变”两组。前者为滑坡组，实验的奖励或难度刺激缓慢变化，且变化微小；后者为突变组，实验的奖励或难度刺激突然变化，变化程度较大。滑坡组和突变组的实验轮数相同，通过比较两组最后一轮不道德行为的差异，来证实滑坡效应的存在。以下是两例实验者操纵法的范例。

任务奖励的滑坡操纵。数字矩阵任务(paper-and-pencil matrix task)由 Mazar 提出。任务材料包括一张含多个数字矩阵的试题卷、一张答卷以及装有一定被试费的信封，被试被告知每解决一个矩阵就会得到一定数额的钱。被试需要在试题卷的每个矩阵内圈出两个加起来等于 10 的数字，最后在答卷上写下答对的矩阵数。实验结束时，要求所有被试将试题卷扔到

箱中，只上交答卷，并自行取出信封里相应的报酬，将剩下的钱放回信封(Mazar, Amir, & Ariely, 2008)。这种措施让被试以为自己的真实表现不会被他人知晓，但实际上，研究者可根据匹配的答卷和试题得知他们是否撒谎。在滑坡效应的研究中，研究者将轮数延长至三轮，三轮中的奖励金额设置为逐渐上升或突然上升。在滑坡组，三轮里每矩阵的奖励分别为 0.25 美元、1 美元以及 2.50 美元。在奖励突变条件下，前两轮无奖励，但第三轮会获得每矩阵 2.50 美元的奖励。若第三轮滑坡组的作弊率显著高于突变组，则证实了滑坡效应的存在(Welsh et al., 2015)。类似的改编还有视觉感知任务(perceptual task)、腐败博弈(the corruption game) (Kobis, van Prooijen, Righetti, & Van Lange, 2017)。

任务难度的滑坡操纵。心理数学任务(the mental math task)要求被试解决十道加减题。研究者告知被试：任务开始时一定要按空格键，否则题目的答案会在几秒后出现在屏幕上。实验中，计算机会自动检查被试是否会为了走捷径而作弊(Hippel, Lakin, & Shakarchi, 2005)。滑坡效应的研究中，滑坡组的题目难度持续上升，突变组的题目难度始终较低，直到最后一轮突然上升。若最后一轮的滑坡组作弊量显著高于突变组，则证实了滑坡效应的存在。

2.2 自发性欺骗法

在此类滑坡效应的研究中，研究者并不进行额外操纵，仅将经典范式的任务轮数延长，观察被试的自发性欺骗情况。目前已有的改编来源于道德领域的经典范式：包括无互动对象的单人掷硬币范式、掷骰子范式，有互动对象的双党任务(the two-party task)等等(Fischbacher & Follmi-Heusi, 2013; Weisel & Shalvi, 2015)。单人的自发性欺骗任务较易理解和操纵，例如多轮次地自行投掷硬币或骰子，不同结果对应不同额度的金钱奖励，以此考察被试撒谎的趋势(Henderson, Thoelen, Nadler, Barraza, & Nave, 2018)。双人互动的自发性欺骗研究法较为复杂，但同样将经典范式操作为多轮即可。下面列举被引用较多的双党任务。

双党任务中有两个角色：顾问和估计者，他们需要配合解决一项估算罐子里的硬币数量的任务。区别在于顾问看到的罐子图片清晰，附有硬币数量范围，而估计者看到的图模糊且无范围。正式实验时，顾问需要根据已有信息为估计者提供金额建议；估计者可选择接受或拒绝，并提交最终的估计量。实际上，任务中的顾问为真被试，估计者为假被试。被试会被随机分配到不道德行为“利己-利他”、“利己-损他”、“损己-利他”的三种条件下。“在利己-利他”条件，被试高估越多，双方的奖励都越多；“利己-损他”条件，被试高估越多，自己的奖励越多，对方的奖励越少；“损己-利他”条件和前者相反。除此以外，所有被试都有一个额外赚钱机会：实验结束后，被试被要求对先前任务中出现过的罐子中的硬币量做准确性估计，并根据准确性得到奖励(Cain, Loewenstein, & Moore, 2005)，由此研究者可判断被试是否在实

验中撒谎。实验主要关注“利己-损他”条件，结果发现在这种“坑同伴，奖励多”情境下，被试说谎的程度会随游戏轮数增加而呈现上升趋势。

实际上，大多以道德经典范式为基础的滑坡效应研究都可以从以上两角度入手。例如，研究者也可以在双党任务中设置滑坡和陡崖情境，作为假被试的顾问提供与实际金额偏离程度逐渐增大或是突然增大的金额值，作为真被试的估计者选择接受或拒绝建议。在不同的模式下，考察被试对他人逐渐高估的建议值的接受程度(Gino et al., 2009)。

3 滑坡效应的成因

在过去几十年间，学者不断探讨滑坡效应的内部成因，目前已取得较好的进展。总结以往研究，滑坡效应的心理成因可分为有限道德视角和适应化视角，两种视角的侧重点不同。有限道德视角下的滑坡效应可概括为理智的丧失，认为滑坡效应是理性减弱的结果；适应化视角偏向于感性，认为滑坡效应是感觉、情感适应的结果。下面将详述滑坡效应的几类心理成因。

3.1 理性——有限道德视角

Simon(1956)曾提出有限理性概念(bounded rationality)，认为人的理性介于完全理性和非理性之间。在面临道德冲突时，个体的理性同样会受到限制进而产生耗竭，即有限道德理论(bounded ethicality) (Banaji, Bazerman, & Dolly, 2003; Rees, Tenbrunsel, & Bazerman, 2019)。有限道德视角下的不道德行为是缺乏理性的、直觉化和自动化的，会导致主体在做出不道德行为时处于无意识和非理性状态。这种状态是导致滑坡效应的核心因素，个体在此过程中并未知觉其行为的道德性质。根据国内外学者研究，有限道德视角认为滑坡效应的成因有：道德推脱、自我损耗、内隐偏见。

道德推脱(moral disengagement)。社会要求人们遵循道德价值观，因此个体需要不断地调节利益和自我道德标准之间的矛盾。在保持积极道德自我概念的同时，允许自己犯一些轻微的道德错误(Moore, Detert, Trevino, Baker, & Mayer, 2012)。换言之，个体倾向于合理化这种轻微的道德错误，使自己在得到好处的同时，又不会被迫消极地更新自我道德概念(Gino et al., 2009; Mazar et al., 2008; Shalvi, Handgraaf, & De Dreu, 2011)。随着时间的推移，越来越多的道德越轨行为被纳入道德的自我概念中(Tenbrunsel & Messick, 2004)。最终，腐败便成为常态(Ashforth et al., 2003)。

这种内在心理过程被称为道德推脱，通过道德推脱，人们会对自身不道德的行为产生动机性遗忘(Shu, Gino, & Bazerman, 2011)。以往研究中，道德推脱能解释人们进行各种不道

德行为的倾向，例如攻击行为(Bandura, Barbaranelli, Caprara, & Pastorelli, 1996)、欺骗(Detert, Trevino, & Sweitzer, 2008)、贪污(Moore, 2008)、说谎(Moore et al., 2012)等等。这些证据表明，随着道德推脱程度的增加，个体允许自己的行动范围增加，这也导致个体特别容易在"不道德的滑坡上继续滑落"(Tenbrunsel et al., 2004)。该过程在实验里得到了证明，道德推脱在滑坡效应中起到中介作用(Welsh et al., 2015)。

自我损耗(ego-depletion)。Baumeister(1998)提出了自我损耗理论。人的心理能量有限，当一段时间内多次经历自我控制来对抗本能和习惯，就会消耗大量的心理能量，进而导致自我控制能力暂时下降，产生自我损耗(Baumeister, Bratslavsky, Muraven, & Tice, 1998)。处于自我损耗状态下的人有较低的自我控制感，更倾向于浅层认知加工，认同与原有概念相一致的信息，忽视与之不一致的信息，更容易做出不道德行为(Fischer, Greitemeyer, & Frey, 2007)。例如，缺乏睡眠的个体自我控制能力降低，导致不道德行为增多(Barnes, Gunia, & Wagner, 2015)。另外，Kouchaki 提出了上午道德效应(The Morning Morality Effect)，认为人们在上午更加道德(Kouchaki & Smith, 2014)。并且下午的不道德行为更多，滑坡效应更加明显(O'Connell & Kouchaki, 2014)。原因是一天中有许多事会消耗我们的心理能量，导致自我控制逐渐减弱(Hofmann, Meindl, Mooijman, & Graham, 2018; Wilkowski et al., 2018)。心理能量在反复的不道德行为之间产生损耗，因此，自我损耗也成为滑坡效应的推动因素之一。个体在接连不断的道德抉择下，心理能力逐渐被消耗，导致理智丧失、自控能力减弱而陷入道德滑坡效应。

内隐偏差(implicit biases)。内隐偏差指个体未意识到自己的行为，在道德判断过程中表现出一种自动化的状态，产生非有意的不道德行为(Kern & Chugh, 2009)。我们的道德决策受到一种处于意识范围外的系统偏差的影响，这种系统偏差使我们思维和决策过程更加全面和整体化(Banaji et al., 2003)。但这也容易使我们的道德决策脱离理性，陷入无意识的、自动决策的状态中(Reynolds, 2006)，这种观点认为滑坡效应是由内隐偏差而导致的自动化认知所助长的。内隐偏差使个体意识不到眼前的不道德事件，呈现出自动化判断状态，是滑坡效应产生的内部原因。

Tenbrunsel 认为内隐偏差的过程类似于数学中的归纳机制(induction mechanism)，首先若 n=1 及 n=m 的命题同时成立，可推导出 n=m+1 时命题也成立。这个过程类似于多米诺效应：只要证明任意一张骨牌倒了，与其相邻的下一张骨牌也会倒下，从而得到"所有的骨牌都会倒下"的结论。Tenbrunsel 将归纳机制视为滑坡效应的内部过程，认为人们会以过去的实践作为评估新实践的基准。如果我们过去的行为没有任何道德问题，那么与之类似、仅稍有变化

的新行为便同样可行。内隐偏差让人难以注意行为之间的差异性，行为易驶入非理性的自动化状态(Tenbrunsel et al., 2004)。

Gino 检验了内隐偏差在滑坡效应中的作用，在被试进行了不道德行为实验之后，要求被试完成一项单词填充任务。由于内隐偏差，处于滑坡状态的人更难以激活与道德标准相关的想法。例如当人们被要求完成“ch___”这个词时，如果他们意识到自己在作弊，那么他们将更有可能填“cheat”而不是“chime”。结果发现在滑坡情境下的被试使用的与欺骗行为相关的单词数显著少于在突变条件下的单词数，考虑时间更短(Gino et al., 2009)。表明内隐偏差使人们难以意识行为的道德性质差异，于是越来越容易做出不道德行为。

3.2 感性——适应化视角

除了理性的缺失，处于滑坡效应中的人通常还伴随着情感反应的降低。为了适应由认知和情绪失调带来的心理负担，自我调节系统通常会不断地调节认知或情绪以取得动态平衡。在这个过程中，人会产生自我认知改变和情绪麻木等状态，这种状态会引发不道德行为的加剧，产生滑坡效应。适应化视角下，滑坡效应有伦理麻木和自我非人化两种心理因素。

伦理麻木(psychic numbing)。伦理麻木可以说是适应的结果。一方面，心理学研究表明，重复暴露会导致脱敏，如反复遭受悲伤或苦难之后，人们的情感会变得麻木，如饱经苦难的人对新的痛苦不那么敏感。在重复地面临道德困境时，由于适应而产生的迟钝会使人参与到潜在更多的不道德行为中，该过程称为伦理麻木(Bok, 1989)。伦理麻木会让处于新情境中的人难以注意行为的伦理性，导致不道德行为越来越多。另一方面，以往研究表明，做出不道德行为通常会产生沮丧或内疚等负面情绪(Kayyal, Pochedly, McCarthy, & Russell, 2015)。即撒谎并不容易，其中伴随的难以忍受的情绪体验会让人们停止不道德行为，做出弥补甚至自我惩罚(Regan, 1971; Yoel, Pizarro, Thomas, & Dan, 2013)。Garrett(2016)认为人们之所以会陷入滑坡效应，可能是情绪适应的结果。滑坡效应过程中的功能性核磁共振成像(functional magnetic resonance imaging, fMRI)扫描发现被试脑中的杏仁核随着时间逐渐变得不活跃，而杏仁核是掌管负面情绪的重要脑区。研究者据此推测重复说谎导致了人们对负面情绪的感知麻木，进而产生滑坡效应(Garrett et al., 2016)。

自我非人化(self-dehumanization)。与人类本质相关的道德理论研究表明，道德准则和道德情感是一种独特的人类财产，它将我们与动物区分开来(Haslam, 2006)。研究者通常将极易违反道德准则，或者极度缺乏道德理性和情感的状态称为“非人化”，最初用于形容战争时残暴的统治者。近期，研究者开始将研究视角转向自我非人化，指个体感到自我缺乏人性的一种状态(Kouchaki, Dobson, Waytz, & Kteily, 2018; Yang, Jin, He, Fan, & Zhu, 2015)。例如

被试在电子游戏中故意攻击其他玩家或排斥其他个体之后，会表现出自我非人化，感觉自己不那么有人性(Bastian, Jetten, & Radke, 2012)。

Kouchaki 等人在八项实验中揭示了不道德行为与自我非人化之间的双向路径，发现两者互为因果。也就是说，人们在不道德行为后会感觉到自己的人性不那么正常，导致自我非人化。而自我非人化的结果反过来又会诱发接下来的不道德行为，产生一种螺旋向下的过程(Kouchaki et al., 2018)，导致不道德行为愈发严重，产生滑坡效应。因此在一定程度上，自我非人化有助于解释不道德行为的滑坡效应。

持自我非人化的人(如罪犯)可能会继续迈上不道德之路，因此强调恢复人类情感的解决方案至关重要。消除自我非人性化会让人们相信自己是完整的个体，这会激励他们去扮演符合道德规则的社会角色。以往研究表明道德规范领域的负面新闻会加剧人们的不道德行为(Yao & Yu, 2016)。若媒体不将重点放在让不道德者纠正错误方面，而是强调这种非人化(例如将资本主义者比喻为吸血鬼)，便可能会导致这些人深化非人性化的观念，进而产生更多的不道德行为。

4 滑坡效应的干预

针对滑坡效应的“小恶逐渐升级成大恶”的外部行为表现，及其理性和感性水平下降的内部心理原因。研究表明，可以从两方面进行有效干预。第一，启动预防定向的动机，控制不道德行为的源头，唤醒后续的理智感和情感。第二，传播积极的描述性社会规范，即使人们短时间内陷入了缺乏理性、无意识、自动化状态的滑坡效应中，也很容易因为良好的群体道德氛围而及时调整不端行为。

4.1 预防定向动机

在道德抉择中，持不同动机的人会做出不同的举动。调节定向理论(regulatory focus theory)认为：持促进定向(promotion focus)动机的人呈冒险性倾向，会积极追求以达到目标；持预防定向(prevention focus)的人呈警惕性倾向，会努力避免失败和错误，更注重履行个人的责任和义务(Higgins, 1987)。预防定向能重新激活自我调控的过程，以免道德推脱等内部因素使自我的道德标准放松。即使是不道德行为微不足道的时候，专于预防的人在做决策时都更加谨慎。这种警惕心可以帮助人们避免或及时脱离缺乏理性、无意识和自动化的状态。

研究表明，启动预防定向动机可以从源头减小滑坡效应。Gino 在四项研究中发现预防定向组的人会避免风险，他们在报告成绩的时候更加谨慎，不道德行为也较少(Gino et al., 2009)。预防定向的操纵通常开始于正式实验前(Li et al., 2011; Welsh et al., 2015)。首先，向

预防定向组呈现与预防有关的词语(例如安全、警惕、预防)；控制条件组呈现中性词。接着，预防定向组的被试需要写下他们最不希望出现的结果，以及为了避免该后果所要采用的策略；控制条件则让被试描述一些工作或生活中的场景。在启动任务的最后几分钟里，要求被试一直感受恐惧的心情，考虑要面临的挑战。结果发现预防定向组的被试的不道德行为显著低于控制组，且未出现滑坡效应。

Bazerman 认为当不道德行为逐渐发展时，滑坡效应会使我们的意识消沉。所以即使是微不足道的违规行为，也要保持警惕，并立即进行调整(Bazerman et al., 2011)。提升预防定向水平有助于让人们时刻维持道德感良好的现状，是控制一开始不道德行为、检测并调整后续行为的有效方法。因此即使行为的不道德程度很小，也要当场纠正并进行警示。这可以帮助人们重新激活因道德推脱等因素而逐渐丧失的理性和负面情绪，使其恢复到正常的水平。

4.2 描述性社会规范

以往有关滑坡效应的研究多集中于个体行为变化层面上，但人并非生活于真空中，其行为很大程度受到群体及社会的影响。在滑坡效应使人陷入非理性、情感麻木和自动化状态的同时，也会放大群体的影响力。此刻产生的“从众效应”是极其关键而又危险的，良好的群体道德规范会及时引导人回归正轨，若未形成良好的群体道德规范，将很容易造成群体大规模的“道德滑坡”。

人们的行为受社会规范信息的影响。当前有两种规范类型：描述性社会规范(descriptive norms)和禁止性社会规范(injunctive norms)(Cialdini, Reno, & Kallgren, 1990)。描述性社会规范传达了大多数人在某情境当中的行为选择信息，指某种情境下的“常态(normal)”行为；禁止性社会规范呈现了当前文化下，多数人对某事件的禁止态度，体现在“某事不可为”的一种道德信念。由于描述性社会规范更易操作，通常成为研究者的首选。研究发现，利用积极的描述性社会规范可以有效地减少不道德行为。Köbis 在一项有关官员腐败的研究中发现“人们参与贿赂的普遍度”的信息会极大地影响被试在后续是否选择贿赂行为。研究者据此指出传达正向的描述性社会规范是减少腐败最有效的手段之一，认为多数人的选择信息能够成为影响被试不道德行为的一种力量 (Köbis et al., 2017)。说明通过传播正向、积极的描述性社会规范，能有效减小滑坡效应，塑造良好的道德环境。

如今，媒体、网络往往是描述性社会规范传播最主要的渠道，企业、学校等内群体是描述性社会规范传播最有效的途径。一方面，加大具有良好道德性质的描述性社会规范的宣传力度，并谨慎报道不良事件，有助于提高公民的道德水平。另一方面，企业或学校等团体要思考自身是否拥有深入人心的道德文化，不道德行为是否有明确的界定标准，杜绝或及时治

理不正风气，以建立优质的描述性道德规范。总之，由于人们倾向于服从社会规范，因此一开始就要设置明确的社会道德规范，明确法规，树立榜样，人们在未来便会保持道德。

5 总结与研究展望

尽管近期在探索滑坡效应的内部机制以及预防措施方面已初见成效，但在实证研究中仍处于初步发展阶段。近期，甚至有学者质疑滑坡效应的存在，并提出了陡崖论(the steep-cliff metaphor)(Kobis et al., 2017)。他指出人们出于保持良好的自我形象的动机，不会屈服于一些小小的利益，以避免反复进行不道德行为所带来的紧张情绪与压力(Teper, Inzlicht, & Page-Gould, 2011)。只有极其诱人的利益才能压垮人们的意志(Francesca & Dan, 2012)，如同突然跳下一个极端不道德的陡崖。该观点打破了人们的传统观念，后续研究可从该矛盾点入手，探索滑坡论和陡崖论各自的适用范围和关键的边界条件。此外，有关滑坡效应的心理和神经机制的研究还有很大的探索空间，可考虑从下述方面切入。

5.1 拓宽滑坡效应的心理过程研究

总体来看，滑坡效应的部分心理机制已得到证实，但未涉及的领域还应继续探索。如滑坡效应的后续趋势、内部原因的实证及其适用范围等等都是可探究方向。未来需聚焦不道德行为的时间延续效应，深化人们对不道德行为在时间层面上变化的认识。具体可以从以下角度开展。

第一，滑坡效应中是否存在某些临界值或节点，在这个节点的两端，人们会做出完全不同的道德选择。例如，Effron 等人发现时间的紧迫性能影响滑坡进程，在测试中，临近提交的时刻，被试作弊的几率大幅上升。而在意料之外的多出来的时间里，被试的作弊率大幅下跌(Effron, Bryan, & Keith, 2015)。那么滑坡效应是否受到时间紧迫性的影响？另外，诺贝尔学奖获得者西蒙曾提出决策中的满意原则：在追求利益的时候，人们往往遵循满意原则，只要达到了自己预期的收益即可，而非追求最大化利益(Simon, 1978)。我们已知滑坡效应大多与金钱利益相挂钩，那么当金钱收益到达了被试的满意值时，滑坡效应是否还会持续存在呢？未来研究可以从时间性质、满意度等角度，采用相关范式探究滑坡效应的临界点问题。

第二，当前在滑坡效应的内因研究中，情绪适应性还停留于论述和推测层面。后续研究可以通过考察个体意识的觉醒程度来间接检查情绪的影响，因为情绪的适应通常伴随个体意识觉醒程度的降低。也可以利用内隐情绪和外显情绪相结合的方式，在任务每轮之间插入易于理解、快速且高效的情绪测试题进行即时检测。例如，内隐情绪的测量方式可选取 Huntsinger 等人(2009)的信息处理风格测试，如果被试将注意力放在整体，则说明当时是中

性或积极情绪状态；如果被试将注意力放在局部，这说明当时是消极情绪状态(Huntsinger, Lun, Sinclair, & Clore, 2009)。外显情绪一般为自我报告，如 Russell 的情绪环状模式表格，该方式的优点在于对被试的情绪变化很敏感，且多次使用不会产生疲劳(Russell, Weiss, & Mendelsohn, 1989)。

第三，一个领域呈滑坡式的不道德行为是否会蔓延到其他领域中？例如，一个逐渐消极怠工的员工可能会做出更多与工作无关的不道德行为，如将办公用品私自拿回家，随地吐痰等等。探讨个体生活中的不道德行为是否会蔓延到其工作和学习中，且呈逐渐加剧的趋势，也是一个很有趣的话题。可以通过在实验中设置两类不相关任务，观察被试在一项任务中的不道德行为是否会在另一项任务中得到延续并加剧。

第四，有关个体层面的研究已取得一定的成果，而人际互动层面的道德滑坡效应研究还有待继续探索。目前关于人际层面的滑坡效应研究使用双党任务，发现人们会越来越容易接受他人的不道德行为(Gino et al., 2009)。但该研究还存在局限性，即身为旁观者的被试并非不道德行为的发起者，不用为行为承担责任。未来研究应排除责任扩散的影响，作进一步改进。群际层面的研究还停留在理论层面(Volokh, 2003)，尚未开展实证。根据人际互动层面的滑坡效应的内涵可知其涉及不道德传染(immoral contagion)(Tapp & Occhipinti, 2016)，并且在传染的过程中，道德水平有逐步降低的倾向。未来研究者可结合不道德传染相关理论，对人际层面尤其是群际层面的滑坡效应研究进行改进和开拓。

5.2 深入滑坡效应的神经机制研究

近年研究者开始着手于滑坡效应的神经机制研究。Garrett 首次采用了 fMRI 考察了滑坡效应的神经机制，开拓了滑坡效应的神经学领域 (Engelmann et al., 2016)。然而该研究的被试数量过少，且 fMRI 的时间分辨率不够高。未来可结合 fMRI 以及时间分辨率更强的 EEG(electroencephalography, EEG)，配合被试的主观报告，来揭示心理和生理的综合变化。继 Garrett 之后，不断有研究者在神经机制方面对滑坡效应进行了跟进(Liu et al., 2019; Tang et al., 2019)，例如 Liu 等人(2019)将计算机建模和 fMRI 相结合，用鼻喷催产素的方式干预滑坡效应。Tang 等人(2019)使用了近红外光谱技术(functional near-infrared spectroscopy, fNIRS)对滑坡效应进行了探测，记录了右侧 dlPFC 和 rTPJ 的脑活动。在已有研究的基础上，未来的滑坡效应的神经机制研究还可以从以下方面推进。

第一，对于滑坡效应的研究可考虑扩展到神经网络互动中。早有研究者提出不应局限于大脑中的某个特定区域去研究不道德行为(Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003)。如背外侧和背内侧前额叶皮质都是参与评估道德程度的重要节点(Baumgartner,

Fischbacher, Feierabend, Lutz, & Fehr, 2009; Dogan et al., 2016)。研究表明，从杏仁核到更广的道德相关网络在不道德情境的适应过程中，信号会逐渐减弱，而这整个网络可能跟滑坡效应有联系(Bartra, Mcguire, & Kable, 2013)。因此，未来应该关注滑坡效应是否会伴随某个神经网络的变化。

第二，5-羟色胺可用于滑坡效应中情绪适应的探究。5-羟色胺又名血清素，是一种抑制性神经递质，被认为对情绪的抑制有关，可降低杏仁核因不道德行为产生的负面情绪(Murphy, Norbury, Sullivan, Cowen, & Harmer, 2009)。在道德滑坡效应中，可以观察 5-羟色胺的分泌是否增多，或者用药物暂时促进 5-羟色胺的传递，观察是否会增强滑坡效应。

第三，由于人们希望保持良好的自我道德形象，因此通常而言欺骗行为会导致身心上的冲突(Eisenberg, 2000)。除了本文归纳的心理内因外，同样会造成被试的某些生理或动作等方面的改变。未来可在实验中利用生理多导仪或视频记录，考察被试的生理变化或表情、身体姿势等变化情况，间接推测被试的心理状况。

第四，通过调节大脑皮层神经元有助于抑制或中断滑坡效应。例如，可以采用经颅直流电刺激(Transcranial direct current stimulation, tDCS)来干预滑坡效应，tDCS 是一种利用电信号的非侵入性大脑神经调控技术。尽管已有研究者使用 tDCS 对滑坡效应的脑机制进行了因果效应的探索，但他们刺激的脑区并不一致，分别为 rTPJ(Noguchi & Oizumi, 2018)和 dlPFC(Marechal, Cohn, Ugazio, & Ruff, 2017)。在已有道德研究的基础上，直接刺激某个脑区的方式并不能囊括所有可能在滑坡效应下发挥作用的区域。如以往研究还表明在经颅直流电刺激下，mPFC 活动增强，同样可以唤醒被试的道德感和情绪反应(Yuan et al., 2017)。因此未来可以先后借助 fMRI 和 tDCS，在同一研究中系统地探讨滑坡效应和某些脑区的因果关系。

参考文献

陈银飞，茅宁. (2011). 变化盲视，滑坡效应与供应商伦理管理的实验研究. *软科学, 25*(9), 125–129.

郑琴琴. (2009). 动态视角下的企业道德滑坡行为研究. *经济理论与经济管理, 11*, 69–75.

Ashforth, B. E., & Anand, V. (2003). The Normalization of corruption in organizations. *Research in Organizational Behavior, 25*(19), 1–52.

Barnes, C. M., Gunia, B. C., & Wagner, D. T. (2015). Sleep and moral awareness. *Journal of Sleep Research, 24*(2), 181–188.

Banaji, M. R., Bazerman, M. H., & Dolly, C. (2003). How ethical are you? *Harvard Business Review, 81*(12), 56–64.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs

on academic functioning. *Child Development, 67*(3), 1206–1222.

Bartra, O., Mcguire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage, 76*(1), 412–427.

Baron, R. A., Zhao, H., & Miao, Q. (2015). Personal motives, moral disengagement, and unethical decisions by entrepreneurs: Cognitive mechanisms on the "slippery slope". *Journal of Business Ethics, 128*(1), 107–118.

Bastian, B., Jetten, J., & Radke, H. R. M. (2012). Cyber-dehumanization: Violent video game play diminishes our humanity. *Journal of Experimental Social Psychology, 48*(2), 486–491.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality & Social Psychology, 74*(5), 1252–1265.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron, 64*(5), 756–770.

Bazerman, M. H., & Tenbrunsel, A. E. (2011). Ethical breakdowns. *Harvard Business Review, 89*(4), 58–65.

Bok, S. (1989). *Secrets: On the ethics of concealment and revelation*. New York: Vintage.

Burger, J. M. (2009). Replicating milgram would people still obey today? *American Psychologist, 64*(1), 1–11.

Cain, Daylian M., Loewenstein, G., & Moore, Don A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies, 34*(1), 1–25.

Chugh, D., Kern, M. C., Zhu, Z., & Lee, S. (2014). Withstanding moral disengagement: Attachment security as an ethical intervention. *Journal of Experimental Social Psychology, 51*(2), 88–93.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality & Social Psychology, 58*(6), 1015–1026.

Courtois, C., & Gendron, Y. (2017). The "Normalization" of deviance: A case study on the process underlying the adoption of deviant behavior. *Auditing a Journal of Practice & Theory, 36*(3), 15–43.

Detert, J. R., Trevino, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology, 93*(2), 374–391.

Dogan, A., Morishima, Y., Heise, F., Tanner, C., Gibson, R., Wagner, A. F., & Tobler, P. N. (2016). Prefrontal connections express individual differences in intrinsic resistance to trading off honesty values against economic benefits. *Scientific Reports*, *6*, 33263.

Effron, D. A., Bryan, C. J., & J Keith, M. (2015). Cheating at the end to avoid regret. *Academy of Management Annual Meeting Proceedings, 109*(3), 395–414.

Engelmann, J. B., & Fehr, E. (2016). The slippery slope of dishonesty. *Nature Neuroscience, 19*(12), 1543–1544.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual review of psychology, 51*, 665–697.

Fischbacher, U., & Follmi-Heusi, F. (2013). Lies in disguisean experimental study on cheating. *Journal of the European Economic Association, 11*(3), 525–547.

Fischer, P., Greitemeyer, T., & Frey, D. (2007). Ego depletion and positive illusions: Does the construction of positivity require regulatory resources? *Personality & social psychology bulletin, 33*(9), 1306–1321.

Francesca, G., & Dan, A. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality & Social Psychology, 102*(3), 445–459.

Ganis, G., Kosslyn, S. M., Stose, S., ., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex, 13*(8), 830–836.

Garrett, N., Lazzaro, S. C., Ariely, D., & Sharot, T. (2016). The brain adapts to dishonesty. *Nature Neuroscience, 19*(12), 1727–1732.

Gino, F., & Bazerman, M. H. (2009). When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behavior. *Journal of Experimental Social Psychology, 45*(4), 708–719.

Gino, F., Norton, M. I., & Ariely, D. (2010). The counterfeit self: The deceptive costs of faking it. *Psychological*

Science, 21(5), 712–720.

Gino, F., Schweitzer, M. E., Mead, N. L., & Dan, A. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior & Human Decision Processes, 115*(2), 191–203.

Henderson, A., Thoelen, G., Nadler, A., Barraza, J., & Nave, G. (2018). Testing the influence of testosterone administration on men's honesty in a large laboratory experiment. *Scientific Reports, 8*, 10.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319–340.

Hippel, W. V., Lakin, J. L., & Shakarchi, R. J. (2005). Individual differences in motivated social cognition: The case of self-serving information processing. *Personality & Social Psychology Bulletin, 31*(10), 1347–1357.

Hofmann, W., Meindl, P., Mooijman, M., & Graham, J. (2018). Morality and self-control: How they are intertwined and where they differ. *Current Directions in Psychological Science, 27*(4), 286–291.

Huntsinger, J. R., Lun, J., Sinclair, S., & Clore, G. L. (2009). Contagion without contact: Anticipatory mood matching in response to affiliative motivation. *Personality and Social Psychology Bulletin, 35*(7), 909–922.

Kayyal, M. H., Pochedly, J., McCarthy, A., & Russell, J. A. (2015). On the limits of the relation of disgust to judgments of immorality. *Frontiers in Psychology, 6*, 951.

Kern, M. C., & Chugh, D. (2009). Bounded ethicality the perils of loss framing. *Psychological Science, 20*(3), 378–384.

Köbis, N. C., van Prooijen, J. W., Righetti, F., & Van Lange, P. A. (2015). "Who doesn't?"--The impact of descriptive norms on corruption. *Plos One, 10*(6), e0131830.

Kobis, N. C., van Prooijen, J. W., Righetti, F., & Van Lange, P. A. M. (2017). The road to bribery and corruption: Slippery slope or steep cliff? *Psychological Science, 28*(3), 297–306.

Kouchaki, M., & Smith, I. H. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological Science, 25*(1), 95–102.

Kouchaki, M., Dobson, K. S. H., Waytz, A., & Kteily, N. S. (2018). The link between self-dehumanization and immoral behavior. *Psychological Science, 29*(8), 1234–1246.

Lafollette, H. (2005). Living on a slippery slope. *Journal of Ethics, 9*(3/4), 475–499.

Li, A., Evans, J., Christian, M. S., Gilliland, S. W., Kausel, E. E., & Stein, J. H. (2011). The effects of managerial regulatory fit priming on reactions to explanations. *Organizational Behavior & Human Decision Processes, 115*(2), 268–282.

Liu, Y., Li, S., Lin, W., Li, W., Yan, X., Wang, X., ... & Ma, Y. (2019). Oxytocin modulates social value representations in the amygdala. *Nature Neuroscience*, *22*(4), 633.

Marechal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with noninvasive brain stimulation. *Proceedings of the National Academy of Sciences of the United States of America, 114*(17), 4360–4364.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633–644.

Milgram, S. (1974). Obedience to Authority: An Experimental View. New York: Tavistock.

Moore, C. (2008). Moral disengagement in processes of organizational corruption. *Journal of Business Ethics, 80*(1), 129–139.

Moore, C., Detert, J. R., Trevino, L. K., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology, 65*(1), 1–48.

Moscovici, S. , & Zavalloni, M. . (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology, 12*(2), 125-135.

Murphy, S. E., Norbury, R., O'Sullivan, U., Cowen, P. J., & Harmer, C. J. (2009). Effect of a single dose of

citalopram on amygdala response to emotional faces. *The British Journal of Psychiatry, 194*(6), 535–540.

Noguchi, Y., & Oizumi, R. (2018). Electric stimulation of the right temporo-parietal junction induces a task-specific effect in deceptive behaviors. *Neuroscience Research, 128*, 33–39.

O'Connell, A., & Kouchaki, M. (2014). In the afternoon, the moral slope gets slipperier. *Harvard Business Review, 92*(5), 34–35.

Regan, J. W. (1971). Guilt, perceived injustice, and altruistic behavior. *Journal of Personality & Social Psychology, 18*(1), 124–132.

Rees, M. R., Tenbrunsel, A. E., & Bazerman, M. H. (2019). Bounded ethicality and ethical fading in negotiations: Understanding unintended unethical behavior. *Academy of Management Perspectives, 33*(1), 26–42.

Reynolds, S. J. (2006). Moral awareness and ethical predispositions: Investigating the role of individual differences in the recognition of moral issues. *Journal of Applied Psychology, 91*(1), 233.

Schauer, F. (1985). Slippery slopes. *Harvard Law Review, 99*(2), 361–383.

Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management, 22*, S16–S27.

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin, 37*(3), 330–349.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493.

Tang, H. H., Zhang, S., Jin, T., Wu, H. Y., Su, S., & Liu, C. (2019). Brain activation and adaptation of deception processing during dyadic face-to-face interaction. *Cortex, 120*, 326–339.

Tapp, C., & Occhipinti, S. (2016). The essence of crime: Contagious transmission from those who have committed moral transgressions. *British Journal of Social Psychology, 55*(4), 756–772.

Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research, 17*(2), 223–236.

Teper, R., Inzlicht, M., & Page-Gould, E. (2011). Are we more moral than we think? Exploring the role of affect in moral behavior and moral forecasting. *Psychological Science, 22*(4), 553.

Volokh, E. (2003). The mechanisms of the slippery slope. *Social Science Electronic Publishing, 116*(4), 1026–1137.

Welsh, D. T., Ordonez, L. D., Snyder, D. G., & Christian, M. S. (2015). The slippery slope: How small ethical transgressions pave the way for larger future transgressions. *Journal of Applied Psychology, 100*(1), 114–127.

Weisel, O., & Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America, 112*(34), 10651–10656.

Wilkowski, B. M., Ferguson, E. L., Williamson, L. Z., & Lappi, S. K. (2018). (How) Does initial self-control undermine later aelf-control in daily life? *Personality and Social Psychology Bulletin, 44*(9), 1315–1329.

Yang, W. Q., Jin, S. H., He, S., Fan, Q., & Zhu, Y. J. (2015). The impact of power on humanity: Self-dehumanization in powerlessness. *Plos One, 10*(5), 19.

Yao, Z., & Yu, R. (2016). The spreading of social energy: How exposure to positive and negative social news affects behavior. *Plos One*, *11*(6), e0156062.

Yin, L. J., & Weber, B. (2019). I lie, why don't you: Neural mechanisms of individual differences in self-serving lying. *Human Brain Mapping, 40*(4), 1101–1113.

Yoel, I., Pizarro, D. A., Thomas, G., & Dan, A. (2013). Moral masochism: On the connection between guilt and self-punishment. *Emotion, 13*(1), 14–18.

Yuan, H., Tabarak, S., Su, W., Liu, Y., Yu, J., & Lei, X. (2017). Transcranial direct current stimulation of the medial prefrontal cortex affects judgments of moral violations. *Frontiers in Psychology*, *8*, 1812.

Psychological mechanism and prevention strategy of slippery slope

OU Honglei[1]; SUN Binghai[1,2]; ZHANG Wenhai[3]; LI Weijian[1]

([1] *College of teacher education, Zhejiang Normal University;* [2] *Tin Ka Ping Moral Education Research Center, Zhejiang Normal University, Jinhua* 321004, *China*)

([3] *Yancheng Institute of Jiangsu Province Mental Health Education Center, Yancheng* 224003, *China*)

Abstract: The slippery slope effect refers to the degree of the individuals' immorality to increase gradually over time, namely, the immoral behavior is repetitive and progressive. It widely exists in the behaviors of the individuals and interpersonal interaction, which will affect the normal social interaction between individuals, and even hinder the harmonious development of society. The research methods of the slippery slope effect include experimenter manipulation and spontaneous deception. There are at least three psychological mechanisms: moral disengagement, implicit biases, and self-dehumanization. The ways to intervene in this effect are promoting preventive orientation motivation and shaping positive descriptive social norms. Future studies should further explore the scope of the slippery slope effect and its neural intervention from the perspective of psychological processes and physiological mechanisms.

Key words: slippery slope; moral disengagement; implicit biases; self-dehumanization; prevention focus; descriptive norms